Microestimates of Multidimensional Child Poverty in West Africa (DSSGx UK 2022)

PROJECT PARTNERS

Save the Children are a non-governmental organisation established to defend and promote the rights and interests of children worldwide. It is membership organisation, made up of Save the Children International and 30 national members that together have 25,000 dedicated staff working across 118 countries. Save the Children responds to major emergencies, delivers innovative development programmes, and ensures children's voices are heard through campaigning and advocacy.

UNICEF – the United Nations Childrens Fund – is a multilateral child-rights organisation established by the UN that works to build a better world for children and families. It works to protect the most vulnerable children worldwide with programmes to provide clean water, food, vaccines, education and protection from violence.

Taken together, Save the Children and UNICEF are involved national policy planning andf advocacy across over 50+ wide-ranging programs that aim to reduce child poverty in developing countries. These programs impact 197 million children each year alone in immunization, sanitation, nutrition and education outcomes.

THE PROBLEM

Child poverty correlates strongly with general economic poverty but has distinct aspects. Experts think of child poverty as a multi-dimensional phenomenon in the sense that children have multiple needs that must be met to flourish. Primary among these needs are adequate access to sanitation, education, nutrition, clean water, healthcare and shelter.   A child may be deprived in some of these dimensions more than others. Different children may be deprived in distinct ways if different needs are not met. Yet all are considered deprived. In order to be effective, policy interventions that aim to alleviate child poverty need to be targeted in a way that match the multifaceted nature of the phenomenon. The data needed to develop such a fine-grained understanding is almost always lacking. As a result, many critical policy and resource allocation decisions are necessarily being made on the basis of very coarse-grained data – often national level statistics. This is a problem because highly aggregated data can mask huge individual and regional inequalities making it difficult to identify

those most in need of assistence. The challenge that Save the Children, jointly with UNICEF, brought to the 2022 DSSGx UK summer programme was to use machine learning to combine multiple geospatial, economic and sociodemographic data sets to predict fine-grained estimates of the distribution of multi-dimensional child poverty across low and middle income countries. If successful, this would allow them to better understand the drivers of child poverty, better allocate limited resources and raise awareness of the inequalities that are masked by national-level statistics.

THE DSSGX UK PROJECT

The data available to the team included public data from the Google Earth Engine on precipitation, elevation, land usage census, etc; Facebook's ads connectivity graph, open data on telecommunication networks, and Facebook's Relative Wealth Index (RWI). The hexagonal tessalation provided by Uber's H3 spatial indexing was used to represent the spatial variation of these data target variables by averaging over the tiles of the tesselation. H3 resolution level 7 was used, meaning that each hexagonal tile had area of approximately 5.2 km^2. Prediction of child poverty, or any other target variable, at this level of resolution from such data requires high resolution ground truth data against which a model can be trained and validated. Survey data from the Demographic and Health Surveys (DHS) program was taken to be this ground truth. This had limitations. In particular, the DHS survey data is patchy and relatively sparse. One consequence of this sparsity was that the available training data did not adequately sample all dimensions of child poverty to allow reliable simultaneous prediction of all dimensions. To get around this problem, it was decided to reduce the number of target variables to two:
- Prevalence: the proportion of children who are below the poverty threshold in at least one of the dimensions of child poverty.
- Severity: the average number of dimensions below threshold for those children that are in poverty.

Taken together, the prevalence and severity are sufficient to make visible the areas with the largest number of the most vulnerable children.

Nigeria was selected as an initial test case. It was chosen because it has a relatively large amount of DHS survey data available for training purposes. The technical task was then to train a machine learning algorithm to predict the prevalance and severity values for each tile in the tesselation given the average values of the input variables (GEE, Facebook ads, telecoms data & RWI) for that tile (and potentially its neighbours.) The algorithm that was selected was LGBM (Light Gradient Boosting Machine), a machine learning framework based on gradient boosted decision trees. LGBM was selected over other approaches due to its emphasis on performance and scalability since the number of tiles for a large country like Nigeria presented a significant computational challenge. Spatial cross validation was used to try to control overfitting. After parameter tuning, the resulting best model validated against state level aggregated data to provide a sense check.

A prototype dashboard was then developed for the majority of nations in a swathe of West Africa stretching Senegal to Nigeria. A sample poverty map produced by the model is shown in Figure 1.
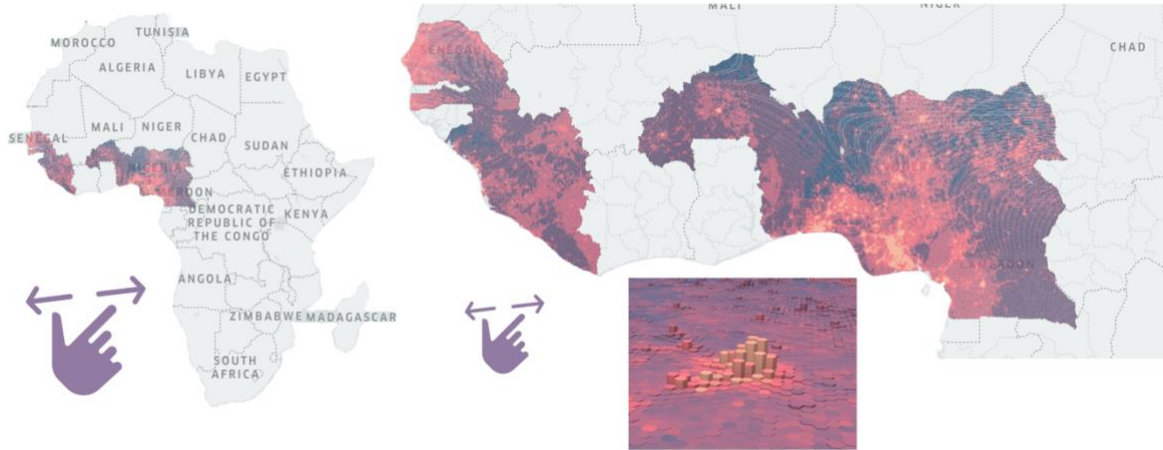
Figure 1: high resolution map of estimated child poverty produced by the DSSG model for a collection of countries in West Africa. The colour map indicates the prevalence, and the height of each tile indicates the severity.

The code from the summer fellowship project can be found here:

- https://github.com/DSSGxUK/s22_savethechildren

A presentation of the project from the 2022 DataFest event in the Shard can be viewed at
- https://youtu.be/MaPjhEjrHgw?t=3

In September 2022, the project was also presented at the Learning and Doing Data for Good conference in Seattle organised by the eScience Institute at the University of Washington. The presentation can be viewed at
- https://youtu.be/R18xqwdpB0s?t=5883

FOLLOW UP WORK

After the completion of the summer fellowship project in August 2022, DSSGx UK fellow, Marina Vicini, chose to continue to develop the tool in collaboration with Oliver Fiala, an analyst at Save the Children's policy and advocacy department. This additional development was done as part of Marina's MSc. in Mathematics for Data Science at the Università degli Studi di Padova. The initial model designed to work for primarily for Nigeria. The team had started looking to expand to other countries as part of the summer fellowship but found some issues that were directly linked to Nigeria which needed working on. During her follow up work, Marina resolved these issues and extended to model to work for all countries in sub-Saharan Africa.

The code for the extended version of the tool is available here:

- https://github.com/marinavicini/stc_continuing

IMPACTS

This project is from the most recent DSSGx UK programme so the impacts are still developing. Save the Children are currently working on a new public platform to upgrade GRID, their existing child inequality tracker (see https://www.savethechildren.net/grid). The purpose of this platform is to bring together aggregated and disaggregated data on key Sustainable Development Goals indicators

related to children's wellbeing, including for child survival and nutrition, child protection, and education and outcome data associated with Save the Children's programmes. The intention is to integrate the child poverty tool developed by DSSGx UK into this platform which is currently scheduled to go live at the end of May. This platform will build an evidence base to help focus the attention of decision makers on the geographical areas that are most affected by child poverty instead of focusing on headline aggregate statistics. Save the Children analyst, Oliver Fiala, indicated that the ability of the tool to make visible the inequalities that are hidden in aggregated national or regional statistics is already turning out to be very helpful in engaging with policymakers, funders, and advocates since it is brings the data much closer to the kinds of questions that they care about and helps them think more clearly about the meaning of aggregated statistics.

The work will become more visible both internally and externally over the coming months. UNICEF is producing a handbook on child poverty which will feature a chapter based on the work done during this project. Save the Children are planning a blog post to accompany the launch of the new GRID platform.

There has also been a lot of positive cross-organisational sharing of the project. For example, Save the Children recently received very complimentary feedback from the World Bank about the work. In April, Oliver Fiala is planning to present the work to the UN Development Programme (UNDP) at a meeting in April with a possible view to working collaboratively on aspects of UNDP work on the Human Development Index.

The model has also opened the door to many follow-on research opportunities. Save the Children have already started doing some specific work exploring the links between climate risk and child poverty. The hope is to be able to quantify at a much more granular level the extent to which poorer areas are exposed to greater climate risks.

> *"High-quality and granular data on child poverty is essential for governments, civil society organisations and many others to better understand and address violations of child rights. But in many cases, we don't know exactly where the poorest children live. The DSSG fellows have done an extraordinary work to build a publicly accessible model which predicts child poverty on a very local level. We are now continuing expanding this analysis to many more countries and are excited using this work in our advocacy and policy work going forward."*
> **Oliver Fiala, Save the Children**